RAPIDMINER'S IMPLEMENTATION IN ANALYZING STUDENT MAINSTREAM DATA

Dewi Anggraeni[1*], Rizaldi[1]
[1]Information Systems,Sekolah Tinggi Manajemen Informatika dan Komputer Royal, Indonesia

Dewi Anggraeni
anggraeni1987@gmail.com

Keywords:
C4.5 algorithm
mainstream course
rapid miner

ABSTRACT

Mainstream is the choice of course to be a guide for students to be attentive in terms of knowing the abilities possessed by each student. C4.5 algorithm method is a method used for predictive classification that aims to find patterns from relatively large to very large data. Rapidminer is a computer application that is used to process data so the data can be useful information. Based on the result of rapidminer implementation in analyzing student mainstream shows the value of achievement index can be the basis for students to know the abilities they have.

INTRODUCTION

Technology affect the attitude and mindset of student development, therefore students must be smart in utilizing technology. Every student has the ability he has, especially in term of education. Based on graduation prediction, that student must have the skills or abilities they have[1]. The word of education has a purpose for improve the intelegence, expertise and skills of students [2]. There are several factors for students to get a job after going through tertiary education, one of the requerements for prospective employees is to have higher tertiary education, as well as their abilities [3]. With students knowing their abilities they will become supporters to complete their thesis faster [4]. The choice of study program must also be considered by students. Thus students can find out how long the study period [5]. One from of parents' attention to children's learning needs and the seriousness of students in the learning process [6].

Various criteria are considered in order to get the best decision, inlcuding influenced by the criteria of interest, talent, cost and the surrounding environment. These various consideration make it difficult for students to make decision, so finally not a few students who choose wrongly in determining the course that suits them [7]. In addition to the study programs that must be considered by students, the selection of majors must be considered, so that academic results are in accordance with the ability of students [8]. The result of the analysis of the C4.5 algorithm for the mainstreaming of mainstream electives shows that the percentage to be considered in the selection of mainstream subjects is 73% [9]. Choosing a major As with product selection, mistakes in product selection can be detrimental to consumer [10]. The purpose of this study is to implement rapidminer in analyzing the mainstream of students showing the value of the

achievement indek can be the basis for student to know the abilities they have. The benefits for students to be taken into consideration in choosing mainstream course.

METHOD

In this study, the initial step carried out by researchers in observation. This research was conducted at the STMIK Royal Kisaran campus. The data used is based on study programs in the form of student grades, student mapping data., course data and mainstream data course. Next, the researchers conducted a study of method. The method of study that researchers do is looking at data on the result of student academic grades, and researchers conduct a literatur review to determine the appropriate method. After that researchers conducted and data testing using rapidminers.
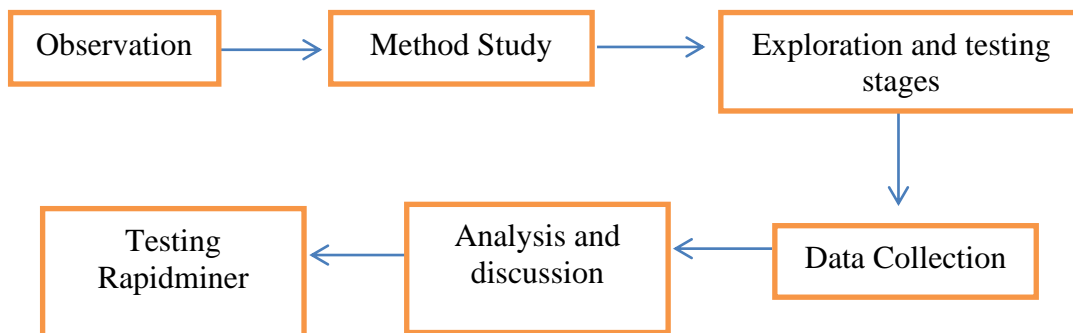

Image 1. Method

RESULT AND DISCUSSION

Knowledge discovery stage in database is a process that is carried out before the data mining analysis stage is carried out. The initial stages of knowledge discovery in database starts from collecting data, using historical data and connecting pattern in large data. Analysis of the data used in this study are subject data, and mainstream subjects. The data is processed first using excel. After that do the analysis method. The method used in this study is the C4.5 algorithm method.

Before implementation rapidminer, the research data were analyzing using excel. At the analysis stage, the method used is the C4.5 algorithm method. The stage for building a decision tree is to prepare training data. After that, calculate the root of the tree. The root will be taken by calculating the highest gain value of each attribute. The highest gain value will be the first root. To calculate the entropy value, the formula used is as follows :

$$Entropy(S) = \sum_{i=1}^{n} Pi \, Log_2 \, Pi \qquad (1)$$

Information :

S = case set

n = number of partision S

Pi = Proportion of Si against S

The calculate the gain value using the following formula :

$$Gain(S,A) = Entropy(S) - \sum_{i=1}^{n} |\frac{Si}{S}| Entropy(Si) \qquad (2)$$

Information :

S = Case set

A = feature

n = number of partition

|Si|= Proportion Si against S

|S|= number of cases in S

Table 1. Entropy Value and Gain Value (Node 1)

attribute	Variable	Number of case	NMP	NDW	NAP	Entropy	Gain
		100	59	30	11	0.814704	
IP Semester 1							
	Tinggi	50	20	20	10	0.691312	0.851483
	Sedang	30	15	10	5	0.774785	
	Kecil	20	10	5	5	0.75	
IP Semester 2							0.762067
	Tinggi	57	20	27	10	0.741499	
	Sedang	43	20	20	3	0.860508	
	Kecil	0	0	0	0	0	
IP Semester 3							0.753142
	Tinggi	58	25	25	8	0.758295	
	Sedang	42	20	20	2	0.900593	
	Kecil	0	0	0	0	0	
Jenis Kelamin							0.921732
	Laki-laki	49	20	20	9	0.706578	
	Perempuan	51	20	27	4	0.888728	

(attribute column: Node 1)

Based on the calculate of entropy and gain values in table 7, the highest gain is found in the IP semester 1 variable with the gain value of 0.851483. Do the same calculation so find the entropy value and the subsequent again. IP atribute variable can be used as root and node 1. Node 1 can be manually described as a decision tree or decision tree as shown below

There are 3 main courses, namely mobile programming, data warehouse, analysis and E-Business design. The research data used in this study are the data values of semester 1 to semester 5 students. Continue the transformation process to classify student value attributes into 3 variabel namely "small" for grades 3, "medium" for grades >3 to <3.5, "large" for values >3.5. Rapidminer is a data processing software that provides tools for making decision trees. The steps taken to implement rapidminer are as follows.

Student grades for semester 1 to semester 5 are grouped according to the attributes to find out the highest gain value. Data processing is performed using the excel application. Next install rapidminer, choose new. Import data that has been processed using excel. Display data import.
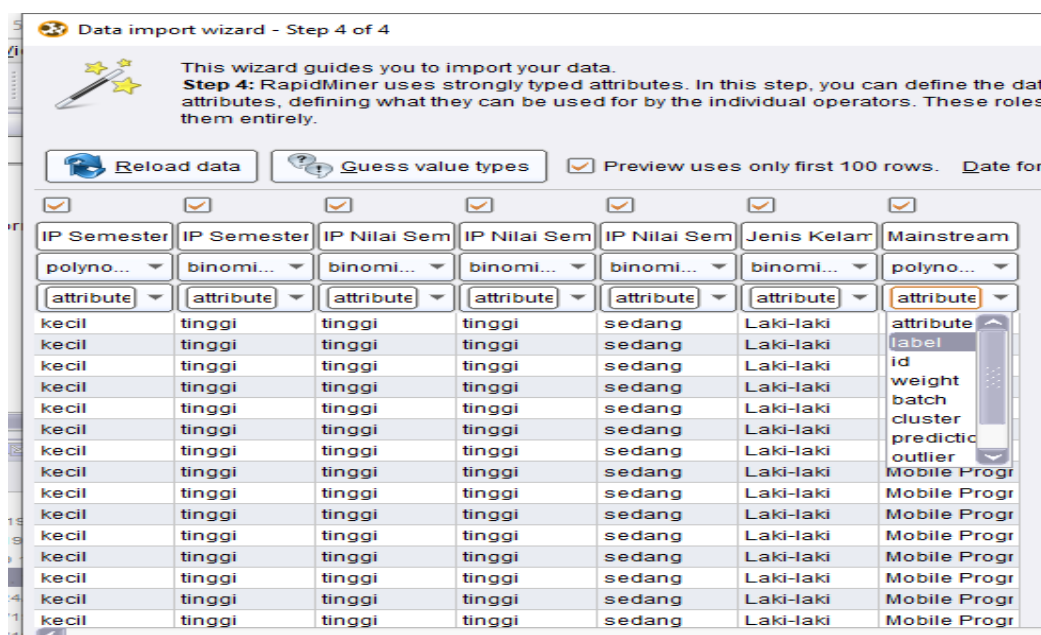

Image 2. Import Data from Excel

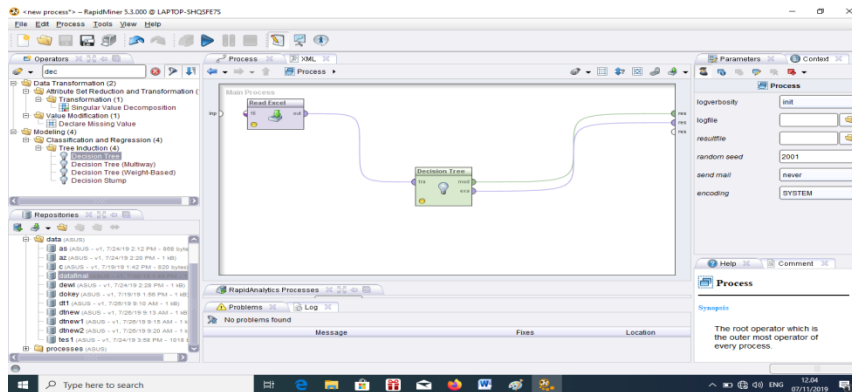Add decision tree operator, select No. pre pruning and pruning.

Image 3. Connection to Decision Tree

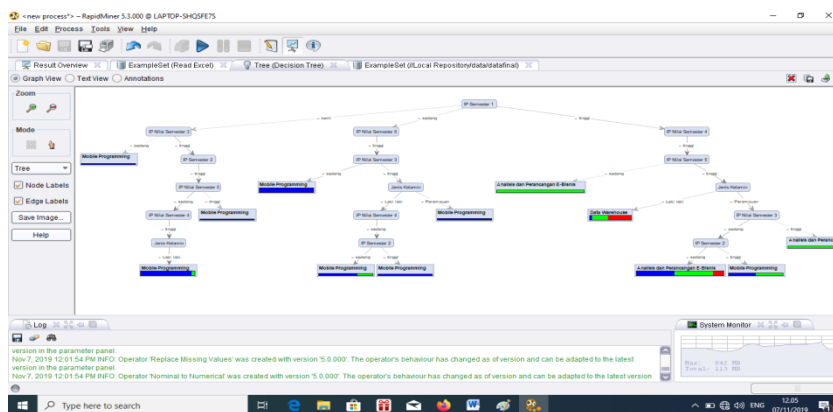The next, Click the run menu to see the result of decision tree.

Image 4. The Result Rapidminer

CONCLUSION

The study program is a unit of educational and learning activities that has a curriculum and learning methods. Each study program has a curriculum, where the curriculum, is a set of planning arrangements to achieve graduate learning outcomes, study materials, processes and assesments that are used a guidelines for organizing study programs. Mainstream elective courses is a choice of selective courses that can become the focus of students in determining interests and talents. C4.5 algorithm is one method of algorithm that has a decision tree, by utilizing data from academics about mainstream selection.

Rapidminer is a computer application that is used to process data so the data can be useful information. Based on the result of rapidminer implementation in analyzing student mainstream shows the value of achievement index can be the basis for students to know the abilities they have.

BIBLIOGRAPHY

[1] M. Mardalius, "PEMANFAATAN RAPID MINER STUDIO 8.2 UNTUK PENGELOMPOKAN DATA PENJUALAN AKSESORIS MENGGUNAKAN ALGORITMA K-MEANS," *JURTEKSI*, 2018, doi: 10.33330/jurteksi.v4i2.36.

[2] Rizaldi and Anggraeni. D, "SISTEM PENUNJANG KEPUTUSAN PENERIMAAN SISWA BARU DENGAN MENGGUNAKAN METODE LOGIKA FUZZY PADA SEKOLAH MENENGAH TINGKAT ATAS (SMA) NEGERI 1 TANJUNG RAYA AGAM," *J. Teknol. dan Sist. Inf.*, vol. 3, no. 1, pp. 1–10, 2016.

[3] F. F. Harryanto, S. Hansun, U. M. Nusantara, G. Serpong, and C. Pegawai, "Penerapan Algoritma C4 . 5 untuk Memprediksi Penerimaan Calon Pegawai Baru di PT WISE," vol. 3, no. 2, pp. 95–103, 2017.

[4] F. Nasari, S. Informasi, P. Keputusan, and D. Selection, "Penerapan algoritma c4.5 dalam pemilihan bidang peminatan program studi sistem informasi di stmik potensi utama medan," pp. 30–34, 2014.

[5] P. Studi, T. Informatika, U. Harapan, and M. Sumatera, "Decision Tree Penentuan Masa Studi Mahasiswa Prodi Teknik Informatika (Studi Kasus : Fakultas Teknik dan Komputer Universitas Harapan Medan)," vol. 5341, no. April, pp. 16–24, 2018.

[6] A. Z. Syah, "Monitoring Orang Tua Terhadap Pembayaran Uang Kuliah Dan Absensi Mahasiswa Berbasis SMS Gateway," vol. 6, no. 4, pp. 374–380, 2019.

[7] R. R. Yuma. Febby Madona, "DI PERGURUAN TINGGI DENGAN METODE PROFILE MATCHING Implementasi Sistem Implementasi User Interface Implementasi user interface dilakukan," vol. 9986, no. September, pp. 3–6, 2018.

[8] L. Swastina, "Penerapan Algoritma C4 . 5 Untuk Penentuan Jurusan Mahasiswa," vol. 2, no. 1, 2013.

[9] D. Anggraeni and T. Christy, "Analisa Kinerja Algoritma C4 . 5 Dalam Menentukan Pola Dominasi Mainstream Mahasiswa," vol. 6, no. 4, pp. 333–339, 2019.

[10] D. Anggraeni, R. Rahim, and A. Z. Syah, "Decision Support System For Formula Milk Selection Based On Nutrition Value Using Tahani Model Database Fuzzy Method Decision Support System For Formula Milk Selection Based On Nutrition Value Using Tahani Model Database Fuzzy Method," 2018.